

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
18 January 2001 (18.01.2001)

PCT

(10) International Publication Number  
WO 01/04147 A2

(51) International Patent Classification: C07K 14/00

(21) International Application Number: PCT/US00/18778

(22) International Filing Date: 11 July 2000 (11.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/143,411 12 July 1999 (12.07.1999) US

(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]: 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

(72) Inventor; and

(75) Inventor/Applicant (for US only): CAHOON, Rebecca, E. [US/US]: 2331 West 18th Street, Wilmington, DE 19806 (US).

(74) Agent: GEIGER, Kathleen, W.: E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/04147 A2

(54) Title: PLANT INOSITOL POLYPHOSPHATE PHOSPHATASE HOMOLOGS

(57) Abstract: This invention relates to an isolated nucleic acid fragment encoding an inositol polyphosphate phosphatase. The invention also relates to the construction of a chimeric gene encoding all or a substantial portion of the inositol polyphosphate phosphatase, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the inositol polyphosphate phosphatase in a transformed host cell.

Shi et al.  
Serial No. 10/042,894

REF  
A3

# TITLE

## PLANT INOSITOL POLYPHOSPHATE PHOSPHATASE HOMOLOGS

This application claims the benefit of U.S. Provisional Application No. 60/143,411, filed July 12, 1999.

### FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding inositol polyphosphate phosphatase in plants and seeds.

### BACKGROUND OF THE INVENTION

*Myo*-inositol 1,2,3,4,5,6-hexaphosphate, commonly known as phytic acid, is an abundant molecule in many plant seeds and vegetative tissue such as roots and tubers (Hartland and Oberlaeas, (1986) *J. Assoc. Off. Anal. Chem.* 69:667-670). Phytic acid exists primarily as a mixture of potassium, calcium, iron, zinc and magnesium phytate salts (Pernollet J. C. (1978) *Phytochemistry* 17:1473-1480).

In corn (*Zea mays* L.), 90% of the phytate is deposited in protein bodies localized in the germ whereas in legume crops 90% of the phytate is localized in the endosperm and cotyledons. Up to 80% of phytate is in the aleurone layer of wheat (*Triticum aestivum* Lam.) and rice (*Oryza sativa* L.) (O'Dell B. L. et al. (1972) *J. Agric. Food Chem.* 20:718-721). The presence of phytate phosphorous in such food crops decreases the bioavailability of zinc by forming a very stable insoluble phytate zinc complex, making the zinc unavailable in the intestinal mucosa of mammals (O'Dell, B. L., et al. (1972) *J. Agr. Food Chem.* 20:718-721). Although phytate phosphorous is readily available to ruminants, it is less available to monogastric animals. In addition to being only partially digestible, the presence of phytic acid in food crops leads to excretion of other limiting nutrients such as essential amino acids, calcium and zinc (Mroz, Z. et al. (1994) *J. Animal Sci.* 72:126-132; Fox et al., In *Nutritional Toxicology* Vol. 3, Academic Press, San Diego (1989) pp. 59-96).

Phytic acid is thought to arise in plants by two pathways. The first pathway uses free *myo*-inositol as the initial substrate, with subsequent phosphorylation by a phosphoinositol kinase. Contribution to the free *myo*-inositol pool is either by recycling from other pathways or by the dephosphorylation of *myo*-inositol-1-phosphate. The alternate pathway uses *myo*-inositol-1-phosphate as the initial substrate, with subsequent phosphorylations catalyzed by phosphoinositol kinase. The committed step for *myo*-inositol-1-phosphate production is the NAD<sup>+</sup>-catalyzed oxidation of carbon 5 of the β-enantiomer of D-glucose-6-phosphate. This reaction is catalyzed by *myo*-inositol-1-phosphate synthase (Raboy, V. In *Inositol Metabolism in Plants* (1990) Wiley-Liss, New York, pp. 55-76).

Phytic acid is degraded in plant cells to D-*myo*-inositol 1,2,4,5,6-pentakisphosphate and orthophosphate through the action of inositol polyphosphate phosphatase. Manipulation of this enzyme activity could lead to a reduction of phytic acid levels in seeds

and an increase in inositol trisphosphate and free phosphate, thus making phosphorus more metabolically available to animals that are fed the seed. Another method to lower phytic acid levels is by inhibiting the activity of *myo*-inositol-1(or 4)-monophosphatase, which catalyzes the reaction: *myo*-inositol 1-phosphate + H<sub>2</sub>O = *myo*-inositol + orthophosphate.

- 5 Manipulation of the activity of this enzyme in developing seeds could decrease phytic acid levels in seeds and increase levels of free phosphate. Lastly, phytic acid levels could also be reduced by inhibiting the activity of inositol trisphosphate kinase. This enzyme catalyzes the reaction: ATP + 1D-*myo*-inositol 1,3,4-trisphosphate = ADP + 1D-*myo*-inositol 1,3,4,6-tetrakisphosphate. This reaction is one of the final steps leading to the formation of
- 10 *Myo*-Inositol 1,2,3,4,5,6-hexaphosphate (phytic acid). Reduction in the activity of the enzyme in developing seeds would interrupt phytic acid synthesis leaving the phosphate as the more metabolically available inositol trisphosphate and free phosphate.

- In the United States, corn accounts for about 80% of the grain fed to all classes of livestock, including poultry, and is usually ground before feeding (Corn: Chemistry and
- 15 Technology, 1987, American Association of Cereal Chemists, Inc., Edited by Stanley A. Watson and Paul E. Ramstad). A meal with decreased amounts of phytic acid and increased amounts of available phosphate would lead to improved feed efficiency in corn-containing rations, making available certain minerals especially zinc, magnesium, iron and calcium. Indeed, enzymatic treatment of soybean meal-containing rations to partially
- 20 hydrolyze the phosphate groups from phytic acid improves both phosphate availability and the availability of other limiting nutrients. Also, in the wet milling of corn, phytate in the steepwater tends to precipitate, causing problems in handling, storing and transportation of the steep liquor. (Pen et al. (1993) *Biotechnology* 11:811-814). In light of these factors, it is apparent that corn plants with heritable, substantially reduced levels of phytic acid and
- 25 increased levels of free phosphorous in their seeds would be desirable. Accordingly, the availability of nucleic acid sequences encoding all or a portion of these enzymes would facilitate studies to better understand carbohydrate metabolism and function in plants, provide genetic tools for the manipulation of these biosynthetic pathways, and provide a means to control carbohydrate transport and distribution in plant cells.

30 SUMMARY OF THE INVENTION

- The present invention concerns an isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a first nucleotide sequence encoding a polypeptide of at least 340 amino acids having at least 80% identity based on the Clustal
- 35 method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 12, 14, 18, 20, 24 and 26; (b) a second nucleotide sequence encoding a polypeptide of at least 97 amino acids having at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of

SEQ ID NOs:8, 10, 16, 22 and 28; and (c) a third nucleotide sequence comprising the complement of (a) or (b).

In a second embodiment, it is preferred that the isolated polynucleotide of the claimed invention comprises a nucleotide sequence which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 or 27 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 or 28.

In a third embodiment, this invention concerns an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 and the complement of such nucleotide sequences.

In a fourth embodiment, this invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to at least one suitable regulatory sequence.

In a fifth embodiment, the present invention concerns a host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast or a plant cell, or prokaryotic, such as a bacterial cell. The present invention also relates to a virus, preferably a baculovirus, comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

In a sixth embodiment, the invention also relates to a process for producing a host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention, the process comprising either transforming or transfecting a compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

In a seventh embodiment, the invention concerns an inositol polyphosphate phosphatase polypeptide selected from the group consisting of: (a) a first polypeptide of at least 340 amino acids comprising at least 80% identity based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 12, 14, 18, 20, 24 and 26; and (b) a second polypeptide of at least 97 amino acids comprising at least 80% identity based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:8, 10, 16, 22 and 28.

In an eighth embodiment, the invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of an inositol polyphosphate phosphatase polypeptide or enzyme activity in a host cell, preferably a plant cell, the method comprising the steps of: (a) constructing an isolated polynucleotide of the present invention or a chimeric gene of the present invention; (b) introducing the isolated polynucleotide or the chimeric gene into a host cell; (c) measuring the level of the inositol polyphosphate phosphatase polypeptide or enzyme activity in the host cell containing the isolated

polynucleotide; and (d) comparing the level of the inositol polyphosphate phosphatase polypeptide or enzyme activity in the host cell containing the isolated polynucleotide with the level of the inositol polyphosphate phosphatase polypeptide or enzyme activity in the host cell that does not contain the isolated polynucleotide.

- 5 In a ninth embodiment, the invention concerns a method of obtaining a nucleic acid fragment encoding a substantial portion of an inositol polyphosphate phosphatase polypeptide, preferably a plant inositol polyphosphate phosphatase polypeptide, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous  
10 nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a substantial portion of an inositol polyphosphate phosphatase amino  
15 acid sequence.

- In a tenth embodiment, this invention relates to a method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding an inositol polyphosphate phosphatase polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA  
20 clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

- In an eleventh embodiment, this invention concerns a composition, such as a hybridization mixture, comprising an isolated polynucleotide or an isolated polypeptide of  
25 the present invention.

- In a twelfth embodiment, this invention concerns a method for positive selection of a transformed cell comprising: (a) transforming a host cell with the chimeric gene of the present invention or a construct of the present invention; and (b) growing the transformed host cell, preferably a plant cell, such as a monocot or a dicot, under conditions which allow  
30 expression of the inositol polyphosphate phosphatase polynucleotide in an amount sufficient to complement a null mutant to provide a positive selection means.

- In a thirteenth embodiment, this invention relates to a method of altering the level of expression of an inositol polyphosphate phosphatase in a host cell comprising:  
35 (a) transforming a host cell with a chimeric gene of the present invention; and (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of the inositol polyphosphate phosphatase in the transformed host cell.

### BRIEF DESCRIPTION OF THE SEQUENCE LISTINGS

The invention can be more fully understood from the following detailed description and the accompanying Sequence Listing which form a part of this application.

- Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825. Table 1 also identifies the cDNA clones as individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), contigs assembled from two or more ESTs ("Contig"), contigs assembled from an FIS and one or more ESTs ("Contig\*"), or sequences encoding the entire protein derived from an FIS, a contig, or an FIS and PCR ("CGS"). Nucleotide sequences, SEQ ID NOs: 3, 5, 7, 11, 13, and 15 and amino acid sequences SEQ ID NOs: 4, 6, 8, 12, 14 and 16 were determined by further sequence analysis of cDNA clones encoding the amino acid sequences set forth in SEQ ID NOs: 18, 20, 22, 24, 26 and 28. Nucleotide SEQ ID NOs: 17, 19, 21, 23, 25 and 27 and amino acid SEQ ID NOs: 18, 20, 22, 24, 26 and 28 were presented in a U.S. Provisional Application No. 60/143,411, filed July 12, 1999.

#### TABLE 1

Inositol Polyphosphate Phosphatase

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Inositol Polyphosphate Phosphatase	etr1c.pk001.i22 (FIS)	1	2
Inositol Polyphosphate Phosphatase	p0121.cfmz67r (CGS)	3	4
Inositol Polyphosphate Phosphatase	rls24.pk0005.f7 (CGS)	5	6
Inositol Polyphosphate Phosphatase	se4.pk0006.d1 (EST)	7	8
Inositol Polyphosphate Phosphatase	srm.pk0010.f10 (FIS)	9	10
Inositol Polyphosphate Phosphatase	wlm96.pk027.m5 (CGS)	11	12
Inositol Polyphosphate Phosphatase	wlm96.pk042.o23 (CGS)	13	14
Inositol Polyphosphate	wre1n.pk0104.b6 (FIS)	15	16

Phosphatase			
Inositol Polyphosphate	Contig composed of:	17	18
Phosphatase	p0121.cfrmz67r		
	p0130.cwttag89r		
	cbn2.pk0011.b9		
	p0105.camag78r		
	p0041.crtbv69r		
	p0056.cddab69r		
Inositol Polyphosphate	rls24.pk0005.f7 (EST)	19	20
Phosphatase			
Inositol Polyphosphate	sc4.pk0006.d1 (EST)	21	22
Phosphatase			
Inositol Polyphosphate	wlm96.pk027.m5 (EST)	23	24
Phosphatase			
Inositol Polyphosphate	wlm96.pk042.o23 (EST)	25	26
Phosphatase			
Inositol Polyphosphate	wre1n.pk0104.b6 (EST)	27	28
Phosphatase			

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical J.* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

#### DETAILED DESCRIPTION OF THE INVENTION

- In the context of this disclosure, a number of terms shall be utilized. The terms "polynucleotide", "polynucleotide sequence", "nucleic acid sequence", and "nucleic acid fragment"/"isolated nucleic acid fragment" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include at least one of 60 contiguous nucleotides, preferably at least one of 40 contiguous nucleotides, most

preferably one of at least 30 contiguous nucleotides derived from SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27 or the complement of such sequences.

The term "isolated polynucleotide" refers to a polynucleotide that is substantially free from other nucleic acid sequences, such as and not limited to other chromosomal and extrachromosomal DNA and RNA. Isolated polynucleotides may be purified from a host cell in which they naturally occur. Conventional nucleic acid purification methods known to skilled artisans may be used to obtain isolated polynucleotides. The term also embraces recombinant polynucleotides and chemically synthesized polynucleotides.

The term "recombinant" means, for example, that a nucleic acid sequence is made by an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated nucleic acids by genetic engineering techniques.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof. The terms "substantially similar" and "corresponding substantially" are used interchangeably herein.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing at least one of 30 contiguous nucleotides derived from the instant nucleic acid

fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by using nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of an inositol polyphosphate phosphatase polypeptide in a host cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide in a virus or in a host cell (eukaryotic, such as plant or yeast, prokaryotic such as bacterial) may comprise the steps of: constructing an isolated polynucleotide of the present invention or a chimeric gene of the present invention; introducing the isolated polynucleotide or the chimeric gene into a host cell; measuring the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide with the level of a polypeptide or enzyme activity in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those

skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms.

- 5 Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the
- 10 temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS which was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

- Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the
- 15 amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85% identical to the amino acid
- 20 sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above identities but typically encode a polypeptide having at
- 25 least 50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal
- 30 method of alignment (Higgins and Sharp (1989) *CABIOS* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

- A "substantial portion" of an amino acid or nucleotide sequence comprises an amino
- 35 acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST

(Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). In general, a sequence of 10 or more contiguous amino acids or 30 or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene.

- 5 Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular
- 10 nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the
- 15 sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

- "Codon degeneracy" refers to divergence in the genetic code permitting variation of
- 20 the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid.
- 25 Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

- "Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art.
- 30 These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to a nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis
- 35 can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of the nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards

those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign gene" refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence. "Regulatory sequences" refers to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or may be composed of different elements derived from different promoters found in nature, or may even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types that are useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory

sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

"Translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

"3' Non-coding sequences" refers to nucleotide sequences located downstream of a coding sequence and includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Engelbrecht et al. (1989) *Plant Cell* 1:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and can be translated into polypeptides by the cell. "cDNA" refers to DNA that is complementary to and derived from an mRNA template. The cDNA can be single-stranded or converted to double stranded form using, for example, the Klenow fragment of DNA polymerase I. "Sense RNA" refers to an RNA transcript that includes the mRNA and can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. "Expression" may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

A "protein" or "polypeptide" is a chain of amino acids arranged in a specific order determined by the coding sequence in a polynucleotide encoding the polypeptide. Each protein or polypeptide has a unique function.

"Altered levels" or "altered expression" refer to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Null mutant" refers to a host cell which either lacks the expression of a certain polypeptide or expresses a polypeptide which is inactive or does not have any detectable expected enzymatic function.

"Mature protein" or the term "mature" when used in describing a protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor protein" or the term "precursor" when used in describing a protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of

methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference). Thus, isolated polynucleotides of the present invention can be incorporated into recombinant constructs, typically DNA constructs, capable of introduction into and replication in a host cell. Such a construct can be a vector that includes a replication system and sequences that are capable of transcription and translation of a polypeptide-encoding sequence in a given host cell. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described in, e.g., Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985, supp. 1987; Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; and Flevin et al., *Plant Molecular Biology Manual*, Kluwer Academic Publishers, 1990. Typically, plant expression vectors include, for example, one or more cloned plant genes under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant expression vectors also can contain a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

"PCR" or "polymerase chain reaction" is well known by those skilled in the art as a technique used for the amplification of specific DNA segments (U.S. Patent Nos. 4,683,195 and 4,800,159).

The present invention concerns an isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) first nucleotide sequence encoding a polypeptide of at least 340 amino acids having at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs: 2, 4, 6, 12, 14, 18, 20, 24 and 26; (b) a second nucleotide sequence encoding a polypeptide of at least 97 amino acids having at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs: 8, 10, 16, 22 and 28; and (c) a third nucleotide sequence comprising a complement of (a) or (b).

Preferably, the nucleotide sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27, that

codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 28.

Nucleic acid fragments encoding at least a substantial portion of several inositol polyphosphate phosphatase enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other inositol polyphosphate phosphatase enzymes, either as cDNAs or genomic DNAs, could be isolated directly by using all or a substantial portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, an entire sequence(s) can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl.*

*Acad. Sci. USA* 86:5673-5677; Loh et al. (1989) *Science* 243:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques* 1:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably one of at least 40, most preferably one of at least 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of an inositol polyphosphate phosphatase polypeptide, preferably a substantial portion of a plant inositol polyphosphate phosphatase polypeptide, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a substantial portion of an inositol polyphosphate phosphatase polypeptide.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing substantial portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1-34; Maniatis).

In another embodiment, this invention concerns viruses and host cells comprising either the chimeric genes of the invention as described herein or an isolated polynucleotide of the invention as described herein. Examples of host cells which can be used to practice the invention include, but are not limited to, yeast, bacteria, and plants.

As was noted above, the nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of phytic acid in those cells.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. The chimeric gene may comprise promoter sequences and translation leader

sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant isolated polynucleotide (or chimeric gene) may be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different cellular compartments, or to facilitate their secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by directing the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.* 100:1627-1632) with or without removing targeting sequences that are already present. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of use may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative

regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of a specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

In another embodiment, the present invention concerns a polypeptide comprising an amino acid sequence selected from the group consisting of (a) a first polypeptide of at least 340 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 12, 14, 18, 20, 24 and 26; and (b) a second polypeptide of at least 97 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:8, 10, 16, 22 and 28.

The instant polypeptides (or substantial portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the

encoded inositol polyphosphate phosphatase. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 6).

All or a substantial portion of the polynucleotides of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and used as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al. In *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping

(Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA* 92:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

#### EXAMPLES

The present invention is further defined in the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

The disclosure of each reference set forth herein is incorporated herein by reference in its entirety.

### EXAMPLE 1

#### Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

- 5 cDNA libraries representing mRNAs from various cattail, corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2  
cDNA Libraries from Cattail, Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
cbn2	Corn developing kernel two days after pollination	cbn2.pk0011.b9
etr1c	Cattail root	etr1c.pk001.i22
p0041	Corn root tips (four days after imbibition), smaller than 5 mm in length	p0041.crtbv69r
p0056	Corn dissected endosperm 13 days after pollination	p0056.cddab69r
p0105	Corn V5** stage roots*	p0105.camag78r
p0121	Corn shank tissue collected from ears 5 days after pollination*	p0121.cfrmz67r
p0130	Corn wildtype internode tissue	p0130.cwtag89r
rls24	Rice leaf (15 days after germination) 24 hours after infection of strain <i>Magaporthe grisea</i> 4360R-67 (avr2-yamo); Susceptible	rls24.pk0005.f7
se4	Soybean embryo, 19 days after flowering	se4.pk0006.d1
wlm96	Wheat seedlings 96 hr after inoculation with <i>Erysiphe graminis f. sp. tritici</i>	wlm96.pk027.m5
		wlm96.pk042.o23
wre1n	Wheat root from 7 day old etiolated seedling*	wre1n.pk0104.b6

- 10 \* These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.  
 \*\* V5 refers to a stage of corn growth. The descriptions can be found in "How a Corn Plant Develops" Special Report No.48, Iowa State University of Science and Technology Cooperative Extension Service Ames, IA, Reprinted  
 15 February 1993.

- cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid  
 20 libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts

will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

#### EXAMPLE 2

##### Identification of cDNA Clones

cDNA clones encoding inositol polyphosphate phosphatase enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet.* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

#### EXAMPLE 3

##### Characterization of cDNA Clones Encoding Inositol Polyphosphate Phosphatase

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to inositol polyphosphate phosphatase from *Arabidopsis thaliana* (NCBI General Identifier No. gi 2160177), *Pichia pastoris* (NCBI General Identifier No. gi 1709725) and *Aspergillus niger* (NCBI General Identifier No. gi 4185610). Shown in Table 3 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), the

sequences of contigs assembled from two or more ESTs ("Contig"), sequences of contigs assembled from an FIS and one or more ESTs ("Contig\*"), or sequences encoding an entire protein derived from an FIS, a contig, or an FIS and PCR ("CGS"):

5

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous to *Arabidopsis thaliana*, *Aspergillus niger* and *Pichia pastoris* Inositol Polyphosphate Phosphatase Proteins

Clone	Status	BLAST pLog Score
Contig composed of:	Contig	
p0121.cfrmz67r		91.30 (gi 2160177)
p0130.cwtag89r		
cbn2.pk0011.b9		
p0105.camag78r		
p0041.crtbv69r		
p0056.cddab69r		
rls24.pk0005.f7	FIS	174.00 (gi 2160177)
se4.pk0006.d1	EST	39.40 (gi 2160177)
wlm96.pk027.m5	FIS	22.70 (gi 1709725)
wlm96.pk042.o23	FIS	23.52 (gi 4185610)
wre1n.pk0104.b6	EST	37.10 (gi 2160177)

10

The sequence of the entire cDNA insert in the clones listed in Table 3 was determined. Further sequencing and searching of the DuPont proprietary database allowed further analysis and characterization of the sequences listed in Table 3 and the identification of other plant clones encoding inositol polyphosphate phosphatase. The BLASTX search using the sequences from clones listed in Table 4 revealed similarity of the polypeptides encoded

15

by the cDNAs to inositol polyphosphate phosphatase from *Arabidopsis thaliana* (NCBI General Identifier No. gi 2160177), *Pichia pastoris* (NCBI General Identifier No. gi 1709725) and *Aspergillus niger* (NCBI General Identifier No. gi 4185610). Shown in Table 4 are the BLAST results for those sequences.

TABLE 4

BLAST Results for Sequences Encoding Polypeptides Homologous to *Arabidopsis thaliana*, *Aspergillus niger* and *Pichia pastoris* Inositol Polyphosphate Phosphatase Proteins

Clone	Status	BLAST pLog Score (NCBI General Identifier No.)
etrlc.pk001.122	FIS	144.00 (gi 2160177)
Contig composed of:	Contig	166.00 (gi 2160177)
cbn2.pk001.1.b9		
p0041.crtbv69r		
p0056.cddab69r		
p0105.camag78r		
p0121.cfrmz67r		
p0130.cwtag89r		
p0121.cfrmz67r	CGS	155.00 (gi 2160177)
rls24.pk0005.f7	CGS	174.00 (gi 2160177)
wlm96.pk027.m5	CGS	22.70 (gi 1709725)
wlm96.pk042.o23	CGS	23.52 (gi 4185610)

The data in Table 5 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14 and 16 and the *Arabidopsis thaliana*, *Aspergillus niger* and *Pichia pastoris* inositol polyphosphate phosphatase sequences.

TABLE 5

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Inositol Polyphosphate Phosphatase

SEQ ID NO.	Percent Identity to (NCBI General Identifier No.)
2	36% (gi 2160177)
4	44% (gi 2160177)
6	44% (gi 2160177)
8	39% (gi 2160177)
10	65% (gi 2160177)
12	20% (gi 1709725)
14	18% (gi 4185610)
16	59% (gi 2160177)

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.,

Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of an inositol polyphosphate phosphatase enzyme.

#### EXAMPLE 4

##### Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axisside facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable

embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and subcultured on this medium every 2 to 3 weeks.

5 The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the napaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

10 The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1  $\mu$ m in diameter) are coated with DNA using the following technique. Ten  $\mu$ g of plasmid DNAs are added to 50  $\mu$ L of a suspension of gold particles (60 mg per mL). Calcium chloride (50  $\mu$ L of a 2.5 M solution) and spermidine free base (20  $\mu$ L of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200  $\mu$ L of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30  $\mu$ L of ethanol. An aliquot (5  $\mu$ L) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

25 For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of mercury (Hg). The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

35 Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the

glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *BioTechnology* 8:833-839).

#### EXAMPLE 5

##### Expression of Chimeric Genes in Dicot Cells

A seed-specific construct composed of the promoter and transcription terminator from the gene encoding the  $\beta$  subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin construct includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire construct is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed construct.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL of liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed construct comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50  $\mu$ L of a 60 mg/mL 1  $\mu$ m gold particle suspension is added (in order): 5  $\mu$ L DNA (1  $\mu$ g/ $\mu$ L), 20  $\mu$ L spermidine (0.1 M), and 50  $\mu$ L  $\text{CaCl}_2$  (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400  $\mu$ L 70% ethanol and resuspended in 40  $\mu$ L of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five  $\mu$ L of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches of mercury (Hg). The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

#### EXAMPLE 6

##### Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoRI and HindIII sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoRI and

Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% low melting agarose gel. Buffer and agarose contain 10 µg/mL ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies, Madison, WI) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs (NEB), Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°C. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

CLAIMS

What is claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

- 5           (a) a first nucleotide sequence encoding a polypeptide of at least 340 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 12, 14, 18, 20, 24 and 26;
- 10           (b) a second nucleotide sequence encoding a polypeptide of at least 97 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:8, 10, 16, 22 and 28; and
- (c) a third nucleotide sequence comprising a complement of (a) or (b).
2. The isolated polynucleotide of Claim 1, wherein the nucleotide sequence of (a) or (b) comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27.
- 15           3. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are DNA.
4. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are
- 20           RNA.
5. A chimeric gene comprising the isolated polynucleotide of Claim 1 operably linked to at least one suitable regulatory sequence.
6. A host cell comprising the chimeric gene of Claim 5.
7. A host cell comprising the isolated polynucleotide of Claim 1.
- 25           8. The host cell of Claim 7 wherein the host cell is selected from the group consisting of yeast, bacteria, and plant.
9. A virus comprising the isolated polynucleotide of Claim 1.
- 10           10. A polypeptide selected from the group consisting of:
- 30           (a) a first polypeptide of at least 340 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 12, 14, 18, 20, 24 and 26; and
- (b) a second polypeptide of at least 97 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a
- 35           polypeptide selected from the group consisting of SEQ ID NOs:8, 10, 16, 22 and 28.

11. A method of selecting an isolated polynucleotide that affects the level of expression of an inositol polyphosphate phosphatase polypeptide in a plant cell, the method comprising the steps of:

- 5 (a) constructing the isolated polynucleotide comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from the isolated polynucleotide of Claim 1;
- (b) introducing the isolated polynucleotide into the plant cell;
- (c) measuring the level of the polypeptide in the plant cell containing the polynucleotide; and
- 10 (d) comparing the level of the polypeptide in the plant cell containing the isolated polynucleotide with the level of the polypeptide in a plant cell that does not contain the isolated polynucleotide.

12. The method of Claim 11 wherein the isolated polynucleotide comprises of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13,  
15 15, 17, 19, 21, 23, 25 and 27.

13. A method of selecting an isolated polynucleotide that affects the level of expression of an inositol polyphosphate phosphatase polypeptide in a plant cell, the method comprising the steps of:

- 20 (a) constructing the isolated polynucleotide of Claim 1;
- (b) introducing the isolated polynucleotide into a plant cell;
- (c) measuring the level of the polypeptide in the plant cell containing the polynucleotide; and
- (d) comparing the level of the polypeptide in the plant cell containing the isolated polynucleotide with the level of the polypeptide in a plant cell  
25 that does not contain the polynucleotide.

14. A method of obtaining a nucleic acid fragment encoding an inositol polyphosphate phosphatase polypeptide comprising the steps of:

- 30 (a) synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 and a complement of such nucleotide sequences; and
- (b) amplifying a nucleic acid sequence using the oligonucleotide primer.

15. A method of obtaining a nucleic acid fragment encoding an inositol polyphosphate phosphatase polypeptide comprising the steps of:

- 35 (a) probing a cDNA or genomic library with an isolated polynucleotide comprising at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID

- NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 and a complement of such nucleotide sequences;
- (b) identifying a DNA clone that hybridizes with the isolated polynucleotide;
- 5 (c) isolating the identified DNA clone; and
- (d) sequencing a cDNA or genomic fragment that comprises the isolated DNA clone.
16. A composition comprising the isolated polynucleotide of Claim 1.
17. A composition comprising the isolated polypeptide of Claim 10.
- 10 18. The isolated polynucleotide of Claim 1 comprising a nucleotide sequence having at least one of 30 contiguous nucleotides.
19. A method for positive selection of a transformed cell comprising:
- (a) transforming a host cell with the chimeric gene of Claim 5; and
- (b) growing the transformed host cell under conditions which allow
- 15 expression of a polynucleotide in an amount sufficient to complement a null mutant to provide a positive selection means.
20. The method of Claim 19 wherein the host cell is a plant.
21. The method of Claim 20 wherein the plant cell is a monocot.
22. The method of Claim 20 wherein the plant cell is a dicot.
- 20 23. A method of altering the level of expression of an inositol polyphosphate phosphatase in a host cell comprising:
- (a) transforming a host cell with the chimeric gene of Claim 5; and
- (b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene
- 25 wherein expression of the chimeric gene results in production of altered levels of an inositol polyphosphate phosphatase in the transformed host cell.
24. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
- (a) a first nucleotide sequence encoding a polypeptide of at least 97 amino
- 30 acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs: 18, 20, 22, 24, 26 and 28; and
- (b) a second nucleotide sequence comprising a complement of the first nucleotide sequence.

## SEQUENCE LISTING

<110> E.I. Du Pont de Nemours and Company

<120> Plant Inositol Polyphosphate Phosphatase Homologs

<130> BB-1377

<140>

<141>

<150> 60/143,411

<151> 1999-07-12

<160> 16

<170> Microsoft Office 97

<210> 1

<211> 1670

<212> DNA

<213> *Typha latifolia*

<400> 1

gcaccagaga	caaccgaaca	aaatgttctc	gcttttttga	gaccattcca	ctggaatcca	60
tagctctctc	tctccgtctc	agagttcaga	gaagcaaacg	aatggcgatc	gccaccatct	120
actccctcct	catacttttc	ctatccgcct	ctctccttcc	ttcgtctcct	tcggtcgagc	180
actcgtttga	tgtgcggcgg	catctttcca	ccgtcacaag	atatgatget	cgtaaagggtg	240
ctgggaatga	cgccctttta	tcttctcttg	taccaaatgg	atgtaggcgg	atccacttaa	300
atcttgtggc	aaggcatgga	accogttctc	ccaccaagaa	acgcatcaaa	gaattggata	360
ggttggaagt	tcgtttgggt	gctcttttaa	aagaagcaaa	acaagagtc	gataaaagt	420
ctctgcagaa	aagtcaccca	tggtagaag	gatgggagtc	accttgga	ggcagggaga	480
aaggtggtag	actagtcagc	aaaggggagg	atgagtata	ccatcttget	attagagtga	540
gggaagggtt	tccagaaata	tttcagagg	agatcaccc	tgaatcttt	actataagg	600
cgaccagggt	tcttcgagcg	tcagctagtg	ctgtagcatt	tggttttggg	ctatttcatg	660
gaaaagggaag	ccttgagacca	ggacaacatc	gcgcttttc	tgtgatcagc	gagagtcgtg	720
caagtgaata	atgtctgcga	tttttgata	ctgtacaac	atacaaggca	tacagacaaa	780
gtcaagagcc	tgtcttgat	aaactcaaa	aaccaatttt	agatgaagtt	acctcttcac	840
tgatctcccg	tcaccaccct	aatttcacaa	ggcaggatgt	cgcttccctt	tggtttctgt	900
cgcaagcagga	agcatctttg	tgggataata	caatcaagc	ttgtggcctt	ttcaatgaaa	960
ctgaggtctc	tttacttgag	tggaccgatg	atttgagggy	ttttctgctt	aaaggttatg	1020
ttaaggcaat	aaattaccgc	atgggcatac	caactactca	agataattgt	caggcaattg	1080
agcaagcgat	catagctaaa	gaagaaaaat	gggcccctcg	aacttttgag	aaagcaagcg	1140
tcgggtttgc	gcatgcagaa	acagttgtgc	ctctcacctg	tcctgttggt	ctttttctcg	1200
aaggatctga	atttgaacag	atacaaaagt	agcaaccact	ttctctact	ccaaagccac	1260
ctcagaagag	aaattggatg	ggcagtatgt	ttgcaccttt	tgctggaaat	aatatatgtg	1320
ctttgtttca	ctgtccgtta	gatgggtctg	atggtgccac	aatgtctaga	gggaaaaata	1380
gcacatacta	tgtgcaagtt	ctacacaagt	aaattccagt	tccaatgcc	ttatcaaat	1440
aacagcactg	tgagaagtga	atttatcttt	ggatattctg	gatacagaat	cttgtgtaac	1500
attgataaat	tacagctctc	ctttaatctc	attcagcatc	cagttaccac	cctgtaggct	1560
gtagttgtgt	gcagctgtgt	aatctaatta	tattatacca	tcaatcatgc	ttcttgattg	1620
agaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa		1670

<210> 2

<211> 446

<212> PRT

<213> *Typha latifolia*

&lt;400&gt; 2

```

Met Ala Ile Ala Thr Ile Tyr Ser Leu Leu Ile Leu Phe Leu Ser Ala
 1           5           10           15

Ser Leu Leu Pro Ser Ser Pro Ser Ala Asp Ser Phe Asp Val Arg
          20           25           30

Arg His Leu Ser Thr Val Thr Arg Tyr Asp Ala Arg Lys Gly Ala Gly
 35           40           45

Asn Asp Ala Leu Leu Ser Ser Ser Val Pro Asn Gly Cys Arg Ala Ile
 50           55           60

His Leu Asn Leu Val Ala Arg His Gly Thr Arg Ser Pro Thr Lys Lys
 65           70           75           80

Arg Ile Lys Glu Leu Asp Arg Leu Glu Ile Arg Leu Gly Ala Leu Leu
          85           90           95

Lys Glu Ala Lys Gln Glu Ser Asp Lys Ser Ser Leu Gln Lys Ser Pro
100           105           110

Pro Trp Leu Glu Gly Trp Glu Ser Pro Trp Lys Gly Arg Glu Lys Gly
115           120           125

Gly Glu Leu Val Ser Lys Gly Glu Asp Glu Leu Tyr His Leu Ala Ile
130           135           140

Arg Val Arg Glu Arg Phe Pro Glu Ile Phe His Glu Glu Tyr His Pro
145           150           155           160

Glu Ile Phe Thr Ile Arg Ala Thr Gln Val Pro Arg Ala Ser Ala Ser
165           170           175

Ala Val Ala Phe Gly Phe Gly Leu Phe His Gly Lys Gly Ser Leu Gly
180           185           190

Pro Gly Gln His Arg Ala Phe Ser Val Ile Ser Glu Ser Arg Ala Ser
195           200           205

Asp Ile Cys Leu Arg Phe Phe Asp Thr Cys Thr Thr Tyr Lys Ala Tyr
210           215           220

Arg Gln Ser Gln Glu Pro Ala Val Asp Lys Leu Lys Glu Pro Ile Leu
225           230           235           240

Asp Glu Val Thr Ser Ser Leu Ile Ser Arg His His Leu Asn Phe Thr
245           250           255

Arg Gln Asp Val Ala Ser Leu Trp Phe Leu Cys Lys Gln Glu Ala Ser
260           265           270

Leu Leu Asp Ile Thr His Gln Ala Cys Gly Leu Phe Asn Glu Thr Glu
275           280           285

Val Ser Leu Leu Glu Trp Thr Asp Asp Leu Glu Gly Phe Leu Leu Lys
290           295           300

Gly Tyr Gly Lys Ala Ile Asn Tyr Arg Met Gly Ile Pro Leu Leu Gln
305           310           315           320

```

Asp Ile Val Gln Ala Met Glu Gln Ala Ile Ile Ala Lys Glu Glu Asn  
 325 330 335  
 Trp Ala Pro Arg Thr Phe Glu Lys Ala Arg Leu Arg Phe Ala His Ala  
 340 345 350  
 Glu Thr Val Val Pro Leu Thr Cys Leu Leu Gly Leu Phe Leu Glu Gly  
 355 360 365  
 Ser Glu Phe Glu Gln Ile Gln Ser Glu Gln Pro Leu Leu Pro Pro  
 370 375 380  
 Lys Pro Pro Gln Lys Arg Asn Trp Met Gly Ser Ile Val Ala Pro Phe  
 385 390 395 400  
 Ala Gly Asn Asn Ile Leu Ala Leu Phe His Cys Pro Leu Asp Gly Ser  
 405 410 415  
 Asp Gly Ala Thr Met Ser Arg Gly Lys Asn Ser Thr Tyr Tyr Val Gln  
 420 425 430  
 Val Leu His Asn Glu Ile Pro Val Pro Met Pro Leu Ser Asn  
 435 440 445

<210> 3  
 <211> 1248  
 <212> DNA  
 <213> Zea mays

<220>  
 <221> unsure  
 <222> (1242)

<400> 3  
 aggttaagtcg agattcgaga aggcaaatcg aattcgccat gggcattggct gctccgcgcg 60  
 cgccgctgcc tctcccccaa ctgctgctcc tctctgttgc cgcgctctcc gccgcgcgctc 120  
 cctcccttag gcgggccagg gcggacgagtg tcgacgtccg ccgccacctc tccaccgtca 180  
 ccaggtatga tctggccagg gattccagta gtgtcatctc catgcgccta atccagaagc 240  
 ggtgcgctgt cattcaactc aatttagtgg caagacatgg gactcgctct cctaccaaga 300  
 agcgcatcaa ggaactggat agattggcag ttgcagtga agccctctgt aaagagcgaa 360  
 atcagctctc tgatagtgat tctctgaaga aaattccatc ctggattaaa ggctgggaat 420  
 cacgctggaa ggttaggact aaaggtgggt agctgattag tgaaggggaa gaggagcttt 480  
 acaatttagc taccagaatg agggagaggt ttcaagatct atttgatgac gaatatcacc 540  
 ctgatgtata ttcaataaga gcaaccaggg ttccctcgagc atcagctagt gcagtggcat 600  
 ttgggttggg actactttct gggaaaggaa agcttggaca agggaagaac cgagcctttt 660  
 ctgtctcgag tgcagtcgt gcaagtgata ttgtctgag attctttgac agctgtgaga 720  
 catcaagggc atacaggaaa aggaaggagc ctgatgtaga gaagcaaaa gaaccaattc 780  
 tagagcatgt cacagctgca ctgtcaatc gttatcacc aaatttaca actccggcatg 840  
 tttcttccct ctgggttctt tgaagcagg aagcatcttt gttgaataga acaaatcaag 900  
 ctgtgggctc tttaaatgaa gctgaggttc gttttctgga gtggacagat gatttggaag 960  
 gttttgtctc aaagggtcat ggtgagtc aaactacag ggaaggagat gccatggctc 1020  
 aaggatgttg tccagtcact ggaagaagca atcatagcta gagaagaaaa ccgtgctgat 1080  
 ggtagctgtg aaaaaggcaag gctccgattt gcacatgca aaactgtgt tcttttagct 1140  
 gccttcttgg tcttttctt gaaggtccag aaattgagaa gatacagaga gaggaagcat 1200  
 tgggacctac cccctttgcc gccacaggga agaaactggg anggcagt 1248

&lt;210&gt; 4

&lt;211&gt; 340

&lt;212&gt; PRT

&lt;213&gt; Zea mays

&lt;400&gt; 4

Met Gly Met Ala Ala Pro Arg Ala Pro Leu Pro Leu Pro Gln Leu Leu  
 1 5 10 15

Leu Leu Leu Val Ala Ala Leu Leu Ala Ala Ala Pro Leu Pro Arg Ala  
 20 25 30

Ala Arg Ala Asp Glu Phe Asp Val Arg Arg His Leu Ser Thr Val Thr  
 35 40 45

Arg Tyr Asp Val Ala Arg Glu Ser Ser Ser Val Ile Ser Met Pro Ser  
 50 55 60

Ile Pro Asp Gly Cys Arg Val Ile His Leu Asn Leu Val Ala Arg His  
 65 70 75 80

Gly Thr Arg Ala Pro Thr Lys Lys Arg Ile Lys Glu Leu Asp Arg Leu  
 85 90 95

Ala Val Arg Leu Glu Ala Leu Leu Lys Glu Ala Asn Gln Val Leu Asp  
 100 105 110

Ser Asp Ser Leu Lys Lys Ile Pro Ser Trp Ile Lys Gly Trp Glu Ser  
 115 120 125

Arg Trp Lys Gly Arg Thr Lys Gly Gly Glu Leu Ile Ser Glu Gly Glu  
 130 135 140

Glu Glu Leu Tyr Asn Leu Ala Thr Arg Met Arg Glu Arg Phe Gln Asp  
 145 150 155 160

Leu Phe Asp Asp Glu Tyr His Pro Asp Val Tyr Ser Ile Arg Ala Thr  
 165 170 175

Gln Val Pro Arg Ala Ser Ala Ser Ala Val Ala Phe Gly Leu Gly Leu  
 180 185 190

Leu Ser Gly Lys Gly Lys Leu Gly Gln Gly Lys Asn Arg Ala Phe Ser  
 195 200 205

Val Leu Ser Glu Ser Arg Ala Ser Asp Ile Cys Leu Arg Phe Phe Asp  
 210 215 220

Ser Cys Glu Thr Tyr Lys Ala Tyr Arg Lys Arg Lys Glu Pro Asp Val  
 225 230 235 240

Glu Lys Gln Lys Glu Pro Ile Leu Glu His Val Thr Ala Ala Leu Val  
 245 250 255

Asn Arg Tyr His Leu Lys Phe Thr Thr Arg Asp Val Ser Ser Leu Trp  
 260 265 270

Phe Leu Cys Lys Gln Glu Ala Ser Leu Leu Asn Thr Thr Asn Gln Ala  
 275 280 285

Cys Gly Leu Phe Asn Glu Ala Glu Val Arg Phe Leu Glu Trp Thr Asp  
 290 295 300

Asp Leu Glu Gly Phe Val Leu Lys Gly Tyr Gly Glu Ser Ile Asn Tyr  
 305 310 315 320

Arg Asp Gly Thr Ala Met Ala Gln Gly Cys Cys Pro Val Asn Gly Arg  
 325 330 335

Ser Asn His Ser  
 340

<210> 5  
 <211> 1928  
 <212> DNA  
 <213> Zea mays

<400> 5  
 ccacgcgtcc ggaaggcaag tcgagattcg aagaaggcaaa tcgaattcgc catgggcgatg 60  
 cgtcgtctcgc gcgcgcgcgt gcctctcccc caactgtctgc tctctctcgt tgcgcgcctc 120  
 ctgcgcgcgcgc ctccccctccc tagggcgggcc agggcggaag agttcgacgt ccgcgcgcac 180  
 ctctccaccgc tcaccaggta tgaatggggc agggagttcca gtatgtcat ctccatccgc 240  
 tcaatccacag acgggtgccg tgcattccac ctcaatttag tggcaagaca tgggactcgc 300  
 gctctctacca agaagcgcac caaggagctg gatagattgg cagttcgact ggagaccctt 360  
 ctgaaagaggg caaatcaggt ccttgatagt gattctctga agaaaattcc atctctgatt 420  
 aaaggctggg aatcacgcgt gaagggtagg actaaaggtg gtgagctgat tagtgaaggg 480  
 gaagaggagc ttacaattt agctaccaga atgagggaga ggtttcaaga tctattttag 540  
 gaggaatatc accctgatgt atattcaata agagcaaccc aggttctcgc agcatcagct 600  
 agtgcagtgg catttgggtt gggactactt tctgggaaag gaaagcttg acaagggaag 660  
 aaccgcagct ttctgttctt gagtgcaggt cgtgcaagtg atatttgtct gagattcttt 720  
 gacagctgtg agacatacaa ggcatacagg aaaagggaag agcctgatgt agagaagcaa 780  
 aaggaaaccaa ttctagagca tgcacagctg gcaactgtca atctgtatca cctaaaaatt 840  
 acaactcgcg atgtttcttc cctctgggtt ctttgtgaagc aggaagcatc tttgttgaat 900  
 acaacaatc aagcttgtgg gctttttaat gaagctgagg tctgtttctt ggagtggaca 960  
 gatgatttgg aggttttgt tctaaaaggc tatggtgagt caattaacta caggatggaa 1020  
 ctgcatttgc tcaaggatgt tgcacagtc atggaagaag caatcatagc tagagaagaa 1080  
 aaccgtgctg atgtgtacgt tgaagaaggca aggcctcgat ttgcacatgc agaaaactggt 1140  
 ttctctttta gctgccttct tggctctttt ctggaaggtc cagaattga gaagatcacg 1200  
 agagagggaag cattgacact accccctttg ccgccacagg gaagaaactg gaaggggcag 1260  
 gttgttgcgc cttttgtcgg taacaatatg ctggttttat acaactctt acttagttca gttctacac 1380  
 tcgtagtgca gcacatactc tggaggccga acaactctt acttagttca gttctacac 1380  
 aacgaagctc cagttttcaat gcctgggttg ggcacaaag atttctgtcc gttcagaggag 1440  
 ttcaaggaga aaattgtgaa accgcacctg aagcagcagt acaacatgat atgcaaggctc 1500  
 aaatccccag cggcaagcga ggagcctgcc tcgttcgcct ccagggtgtc cagtttcttc 1560  
 ctaggactcc tctcgacgaa aggttaccgc ggtgtgggag ccgaggcggt caagaccgag 1620  
 ctgtaggcgt caagacatag cttagcacgg tgcctctcct tctcaccgtc tgcgagcacg 1680  
 atcaggaaac agactaaaccc atgcactaac cgcagctggt ttgtacggg cggagtcgaa 1740  
 tgtgtaggac accaccacag gaggcgccag ttcggttgtt tccagggccca gacgcctctg 1800  
 atactgtagt agcaataaca taaataataa atgtgtttcg atctcgacta aaaaaaaa 1860  
 aaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa aaaggggcgc cgctctagag 1920  
 gatctggt 1928

<210> 6  
 <211> 524  
 <212> PRT  
 <213> Zea mays

<400> 6  
 Met Gly Met Ala Ala Pro Arg Ala Pro Leu Pro Leu Pro Gln Leu Leu  
 1 5 10 15  
 Leu Leu Leu Val Ala Ala Leu Leu Ala Ala Pro Leu Pro Arg Ala  
 20 25 30  
 Ala Arg Ala Asp Glu Phe Asp Val Arg Arg His Leu Ser Thr Val Thr  
 35 40 45  
 Arg Tyr Asp Val Ala Arg Glu Ser Ser Ser Val Ile Ser Met Pro Ser  
 50 55 60  
 Ile Pro Asp Gly Cys Arg Val Ile His Leu Asn Leu Val Ala Arg His  
 65 70 75 80  
 Gly Thr Arg Ala Pro Thr Lys Lys Arg Ile Lys Glu Leu Asp Arg Leu  
 85 90 95  
 Ala Val Arg Leu Glu Ala Leu Leu Lys Glu Ala Asn Gln Val Leu Asp  
 100 105 110  
 Ser Asp Ser Leu Lys Lys Ile Pro Ser Trp Ile Lys Gly Trp Glu Ser  
 115 120 125  
 Arg Trp Lys Gly Arg Thr Lys Gly Gly Glu Leu Ile Ser Glu Gly Glu  
 130 135 140  
 Glu Glu Leu Tyr Asn Leu Ala Thr Arg Met Arg Glu Arg Phe Gln Asp  
 145 150 155 160  
 Leu Phe Asp Asp Glu Tyr His Pro Asp Val Tyr Ser Ile Arg Ala Thr  
 165 170 175  
 Gln Val Pro Arg Ala Ser Ala Ser Ala Val Ala Phe Gly Leu Gly Leu  
 180 185 190  
 Leu Ser Gly Lys Gly Lys Leu Gly Gln Gly Lys Asn Arg Ala Phe Ser  
 195 200 205  
 Val Leu Ser Glu Ser Arg Ala Ser Asp Ile Cys Leu Arg Phe Phe Asp  
 210 215 220  
 Ser Cys Glu Thr Tyr Lys Ala Tyr Arg Lys Arg Lys Glu Pro Asp Val  
 225 230 235 240  
 Glu Lys Gln Lys Glu Pro Ile Leu Glu His Val Thr Ala Ala Leu Val  
 245 250 255  
 Asn Arg Tyr His Leu Lys Phe Thr Thr Arg Asp Val Ser Ser Leu Trp  
 260 265 270  
 Phe Leu Cys Lys Gln Glu Ala Ser Leu Leu Asn Thr Thr Asn Gln Ala  
 275 280 285  
 Cys Gly Leu Phe Asn Glu Ala Glu Val Arg Phe Leu Glu Trp Thr Asp  
 290 295 300  
 Asp Leu Glu Gly Phe Val Leu Lys Gly Tyr Gly Glu Ser Ile Asn Tyr  
 305 310 315 320

Arg Met Gly Leu Pro Leu Leu Lys Asp Val Val Gln Ser Met Glu Glu  
 325 330 335  
 Ala Ile Ile Ala Arg Glu Glu Asn Arg Ala Asp Gly Thr Phe Glu Lys  
 340 345 350  
 Ala Arg Leu Arg Phe Ala His Ala Glu Thr Val Val Pro Phe Ser Cys  
 355 360 365  
 Leu Leu Gly Leu Phe Leu Glu Gly Pro Glu Ile Glu Lys Ile Gln Arg  
 370 375 380  
 Glu Glu Ala Leu Asp Leu Pro Pro Leu Pro Pro Gln Gly Arg Asn Trp  
 385 390 395 400  
 Lys Gly Ser Val Val Ala Pro Phe Ala Gly Asn Asn Met Leu Val Leu  
 405 410 415  
 Tyr Gln Cys Pro Ser Lys Ile Ser Asp Gly Ser Thr Ile Ser Gly Gly  
 420 425 430  
 Arg Asn Asn Ser Tyr Leu Val Gln Val Leu His Asn Glu Val Pro Val  
 435 440 445  
 Ser Met Pro Gly Cys Gly Asn Lys Asp Phe Cys Pro Phe Glu Glu Phe  
 450 455 460  
 Lys Glu Lys Ile Val Lys Pro His Leu Lys His Asp Tyr Asn Met Ile  
 465 470 475 480  
 Cys Lys Val Lys Ser Pro Ala Ala Ser Glu Glu Pro Ala Ser Phe Ala  
 485 490 495  
 Ser Arg Val Ser Ser Phe Phe Leu Gly Leu Ser Gln Lys Gly Tyr  
 500 505 510  
 Arg Gly Val Gly Ala Glu Gly Val Lys Thr Glu Leu  
 515 520

<210> 7  
 <211> 2054  
 <212> DNA  
 <213> Oryza sativa

<400> 7  
 gcacagaggtc aaaaaagctc ccaaaatcaa attaaattaa agagaaaaaa aagaagaaga 60  
 gaaggggcaaa tcgattcgcc atggctgctc ccgcacgccc tctcccctc gtctctctcc 120  
 cgtctctccgc cgtcctctct ccggcggtcc cctctctgcc ggccggccag accgtgtccg 180  
 ccgcgttcga cgtccgcgcg cactctctca ccgtaccag gtacgatgtg gcgaggggat 240  
 ccaatagtgt gtccctccgc ccgtctatgt cggatgagtg ccgcgtgatc cactcaatc 300  
 tcgtggcaag acatgggact cgcgcacctc ccaaaaagag aagggcctgaa agtgactccc 360  
 tggcgggttcg gttgaaggct cttatcgatg aagcaaaaaca agggcctgaa agtgactccc 420  
 tgaaaaaaat tccttcattg atgaaagggg gggagtcacc ctggaaagggt aggggtgaaag 480  
 gtggtgagct ggtcagtgaa ggggaggaag agctatacaa ccttgctatc agagtcgaag 540  
 agaggtttca aggcctattt gatgaggaat atcaccctga tbtgtattca ataaagacaa 600  
 ctcaggttcc tcgggcattca gctagtgcag tagcatttgg ttgggttcta ctttctggga 660  
 aagggaaagt tggacctgtg aaaaaccgtg ccttttctgt tctgagtgag agtgcgtcaa 720  
 gtgatatttg tctcgatttc ttgtatagct gtgaacata caaggactac agggaaaaga 780

```

aggagcctga tgttgaaaag caaaaggaac caattttaga acacgtcaca tcggcattag 840
ttaacgcgtta tcatctcaat tttaacacaa aagatgtttc ttccctctgg ttcctttgca 900
agcaggaagcg atcttttaatg aatataacca atcaagcttg tcaacttttt aatgaagctg 960
agggttaattt tctagatggg acagatgatc tggagggttt tgtgtcaaaa gggttatggg 1020
agtcacaataa ctatcggatg ggaactgccat tgcacaagga cgttgccag tcaatggaag 1080
aagcaatcgt tgcataagaa gaaaaccacc ctgatggtag atatgagaag gcaaggtctc 1140
gatttgacaa tgcagaact gttgtccctt tctcatgtct tcttggtctt ttcttggaag 1200
gtgcagattt tgcgaagata caacgggagg aatcattgga catatcctct gtgccaccac 1260
agggagagaaa ttggaagggc agtgtttgtg cactttttgc tggtaacaat atgttggtct 1320
tgtaccagtg cccaggaaaa actgatgttg gtaagatttc tcgggagtag aagagctcat 1380
acttcgtgca ggttatcac aatgaagctc cagtttcaat gcgggagtc gggaacaaa 1440
atttcgtccc attgaagag ttcaagaga agatagttag accccactg aagcatgact 1500
acgacgcctc atgcaagata aggcgggtg caagagagga gccttcctcc ttacgttcca 1560
ggatgtccaa ttcttccta gttgtttct cgcagaaaag atacctgtt agtgcacagg 1620
atgtgaagtc ggagctgtag gttatacta ggcagatggg gcatcttcta cctttcgac 1680
taagctgcta ctatccatct accatggata accaaggagc gaactcgtg actaaactac 1740
gggtgtttct taactgcaac cctcagtttg atgagcgag ccaatgtga acaacacgct 1800
accactgggt cccctcttct ttccagataga agacgttttg tgatcaatg atctttgtt 1860
ttaactcgga cttctcttag tgcacaactc ctgactctgt tgagatcaaa ctgaaagtga 1920
tagcacacgg ttgtgattgt ggccttctgt gtaggttgag atcatgtact actacttga 1980
caatcgttt caagctttca gctcaatcaa tccaaaaaaa aaaaaaaa 2040
aaaaaaaaaaaa aaaa 2054

```

&lt;210&gt; 8

&lt;211&gt; 519

&lt;212&gt; PRT

<213> *Oryza sativa*

&lt;400&gt; 8

```

Met Ala Ala Pro Arg Thr Pro Leu Pro Leu Val Leu Leu Leu Val Ser
  1                      5                      10                      15
Ala Ala Leu Leu Ala Ala Ala Pro Leu Ser Pro Ala Ala Glu Thr Gly
  20                      25                      30
Ala Ala Ala Phe Asp Val Arg Arg His Leu Ser Thr Val Thr Arg Tyr
  35                      40                      45
Asp Val Ala Arg Gly Ser Asn Ser Val Ser Ser Ala Pro Ser Met Ser
  50                      55                      60
Asp Glu Cys Arg Val Ile His Leu Asn Leu Val Ala Arg His Gly Thr
  65                      70                      75                      80
Arg Ala Pro Thr Lys Lys Arg Ile Lys Glu Leu Asp Arg Leu Ala Val
  85                      90                      95
Arg Leu Lys Ala Leu Ile Asp Glu Ala Lys Gln Gly Pro Glu Ser Asp
  100                     105                     110
Ser Leu Lys Lys Ile Pro Ser Trp Met Lys Gly Trp Glu Ser Pro Trp
  115                     120                     125
Lys Gly Arg Val Lys Gly Glu Leu Val Ser Glu Gly Glu Glu Glu
  130                     135                     140
Leu Tyr Asn Leu Ala Ile Arg Val Lys Glu Arg Phe Gln Gly Leu Phe
  145                     150                     155                     160

```

Asp Glu Glu Tyr His Pro Asp Val Tyr Ser Ile Arg Ala Thr Gln Val  
 165 170 175  
 Pro Arg Ala Ser Ala Ser Ala Val Ala Phe Gly Leu Gly Leu Leu Ser  
 180 185 190  
 Gly Lys Gly Lys Leu Gly Pro Val Lys Asn Arg Ala Phe Ser Val Leu  
 195 200 205  
 Ser Glu Ser Arg Ala Ser Asp Ile Cys Leu Arg Phe Phe Asp Ser Cys  
 210 215 220  
 Glu Thr Tyr Lys Asp Tyr Arg Lys Arg Lys Glu Pro Asp Val Glu Lys  
 225 230 235 240  
 Gln Lys Glu Pro Ile Leu Glu His Val Thr Ser Ala Leu Val Asn Arg  
 245 250 255  
 Tyr His Leu Asn Phe Thr Pro Lys Asp Val Ser Ser Leu Trp Phe Leu  
 260 265 270  
 Cys Lys Gln Glu Ala Ser Leu Met Asn Ile Thr Asn Gln Ala Cys Gln  
 275 280 285  
 Leu Phe Asn Glu Ala Glu Val Tyr Phe Leu Glu Trp Thr Asp Asp Leu  
 290 295 300  
 Glu Gly Phe Val Leu Lys Gly Tyr Gly Glu Ser Ile Asn Tyr Arg Met  
 305 310 315 320  
 Gly Leu Pro Leu Leu Lys Asp Val Val Gln Ser Met Glu Glu Ala Ile  
 325 330 335  
 Val Ala Lys Glu Glu Asn His Pro Asp Gly Thr Tyr Glu Lys Ala Arg  
 340 345 350  
 Leu Arg Phe Ala His Ala Glu Thr Val Val Pro Phe Ser Cys Leu Leu  
 355 360 365  
 Gly Leu Phe Leu Glu Gly Ser Asp Phe Ala Lys Ile Gln Arg Glu Glu  
 370 375 380  
 Ser Leu Asp Ile Pro Pro Val Pro Pro Gln Gly Arg Asn Trp Lys Gly  
 385 390 395 400  
 Ser Val Val Ala Pro Phe Ala Gly Asn Asn Met Leu Ala Leu Tyr Gln  
 405 410 415  
 Cys Pro Gly Lys Thr Asp Gly Gly Lys Ile Ser Arg Asp Gln Lys Ser  
 420 425 430  
 Ser Tyr Phe Val Gln Val Ile His Asn Glu Ala Pro Val Ser Met Pro  
 435 440 445  
 Gly Cys Gly Asn Lys Asp Phe Cys Pro Phe Glu Glu Phe Lys Glu Lys  
 450 455 460  
 Ile Val Glu Pro His Leu Lys His Asp Tyr Asp Ala Leu Cys Lys Ile  
 465 470 475 480

Arg Pro Val Ala Arg Glu Glu Pro Ser Ser Phe Ser Ser Arg Met Ser  
485 490 495

Asn Phe Phe Leu Gly Leu Phe Ser Gln Lys Gly Tyr Arg Val Ser Ala  
500 505 510

Gln Asp Val Lys Ser Glu Leu  
515

<210> 9

<211> 357

<212> DNA

<213> Glycine max

<220>

<221> unsure

<222> (270)

<220>

<221> unsure

<222> (307)

<220>

<221> unsure

<222> (347)

<220>

<221> unsure

<222> (349)

<400> 9

ggaataccat	ccagacatat	atcctattaa	ggcaactcag	gttccccggg	catctgctag	60
tgctgttgca	tttggaaatgg	ggcttttcag	tggaactgga	agtcttgac	ttgggcatca	120
ccgagccctt	gctgttacta	gtgaaagtgc	tgctagcgac	attgtgctga	gatttcacga	180
ttgttgtcat	aattacaagg	ctcatcgga	aagccaggaa	ctgcagttta	gtaaacttaa	240
ggaacctata	ttggatgaga	ttacatctgn	cattgattgg	ggcgccatgg	gctgaatttt	300
acgagcnagg	atacatcctt	ctctctggtt	ttgtgtaaac	aggaagnanc	cttggttg	357

<210> 10

<211> 97

<212> PRT

<213> Glycine max

<220>

<221> UNSURE

<222> (90)

<400> 10

Glu	Tyr	His	Pro	Asp	Ile	Tyr	Pro	Ile	Lys	Ala	Thr	Gln	Val	Pro	Arg
1					5				10					15	

Ala	Ser	Ala	Ser	Ala	Val	Ala	Phe	Gly	Met	Gly	Leu	Phe	Ser	Gly	Asn
		20						25					30		

Gly	Ser	Leu	Gly	Leu	Gly	His	His	Arg	Ala	Phe	Ala	Val	Thr	Ser	Glu
		35						40					45		

Ser Arg Ala Ser Asp Ile Val Leu Arg Phe His Asp Cys Cys His Asn  
50 55 60

Tyr Lys Ala His Arg Lys Ser Gln Glu Pro Ala Val Ser Lys Leu Lys  
65 70 75 80

Glu Pro Ile Leu Asp Glu Ile Thr Ser Xaa Ile Asp Trp Gly Ala Met  
85 90 95

Gly

<210> 11  
<211> 1579  
<212> DNA  
<213> Triticum aestivum

<400> 11  
ctctctgctca atctcaaacg atgtttatca gcagccctct gtctctcgca gtctctctct 60  
ccggtctcata tattgtctcat gcctctatca taaatcagtt tgatccgctg aagcatctat 120  
ctggaggttag tccaccattt gaccacgaag aagccagctc tcctctagac cctgtctccac 180  
ctctaggtctg caatgtcact cgtgcacgat acctggtagc ccatgtctgc atctcaatcca 240  
atagtattga ctacgtagcc ttatcgaac cattctctga aaaaacttag cgacgacccg 300  
ctcagatggcg cgaatccctt agcttatctt ttctagctac ctggagaaac ccaattctaa 360  
agggagagaaa ggaataatta agtcgatccg gtcagctaca agccatgaca ttgtgtgttg 420  
aagtgggaca tagatacttt aatttgagga caccocagaa aatatgggca gcatctctcg 480  
atcgaaagat gaaaagcgct caattctctg caaaaggaat tgctctagac gcatcaaaaga 540  
tagctattga gcgagtgtat gagggcaaga aagatgggtg aaactcaact aatccatatg 600  
atagctgtcc agccttcagt cgagtaactg gcagcgatga agactcaagt ttctcgagga 660  
ctcacacaaa accggttatg caacgcttca actctctggc acctctcttt aatttcaact 720  
ctacggacat ttatgccttt tcgctgattt gcggctacga aacagttctg cgaggttctt 780  
ctccattttg tggattgtcg gtattgagct caaatgaatg gctcggcttt gagtatgcga 840  
atgatatcaa atatttttat aacagtggtt atggcttacc ctacgtgggt gctactggat 900  
ttccatgggt aaatgcaagc tttaatgcgc tcatgtcaaa tcatgatagc aacagtaacg 960  
aggttaaaga tcaagatctg ttctgtatctt ttacacatcg tggatgctt ccaatggcac 1020  
tagttgccat ggggtctctt aataactcag catactcggt agcaataat ataaacgaca 1080  
ccatgcctct cgatactatt aactaccaac gagtctggaa atcaagttag attattccgt 1140  
ttatgactga cattgttatt gaaaaactag aatgtgatag ttctcgcttt gacgaaggga 1200  
tctactatcg cgtcttagta aatgatagac cacagagtct gattggatgt cagctggacg 1260  
cggcgacagc cftgaanaag gaatccattc gcaaatggct aatatgaacg gccaaagtctg 1320  
tcggtgattt tgatactatg tgccgagtag atacgctaa tagcacgaac atattgagca 1380  
tttaccgatt ctaatgatcc tccgccaacg tttttgtata tatgatatag agaacctaa 1440  
atgaattatt cagggcagca ttatttccat cgaagatagg cgttaatgta acttggaagg 1500  
gcattgtctc ttcgattatc ccttaaaaaa aaaaaaaaaa aaactcgagc gggggccgtg 1560  
ccaaatctcc cccccggg

<210> 12  
<211> 457  
<212> PRT  
<213> Triticum aestivum

<400> 12  
Met Phe Ile Ser Ser Pro Leu Ser Leu Ala Leu Leu Ser Gly Ser  
1 5 10 15  
Tyr Ile Ala His Ala Ser Ile Ile Asn Gln Phe Asp Pro Leu Lys His  
20 25 30

Leu Ser Gly Val Ser Pro Pro Phe Asp His Glu Glu Ala Ser Ser Pro  
 35 40 45  
 Leu Asp Pro Val Pro Pro Leu Gly Cys Asn Val Thr Arg Ala Ala Tyr  
 50 55 60  
 Leu Val Arg His Ala Ala Ile His Ser Asn Ser Tyr Asp Tyr Val Ala  
 65 70 75 80  
 Phe Ile Glu Pro Phe Leu Glu Lys Leu Ser Arg Thr Thr Val Arg Trp  
 85 90 95  
 Ala Glu Ile Pro Ser Leu Ser Phe Leu Ala Thr Trp Arg Asn Pro Ile  
 100 105 110  
 Leu Lys Gly Glu Lys Glu Lys Leu Ser Arg Ser Gly Gln Leu Gln Ala  
 115 120 125  
 Met Thr Phe Gly Val Glu Val Gly His Arg Tyr Phe Asn Leu Arg Thr  
 130 135 140  
 Pro Gln Lys Ile Trp Ala Ala Ser Ser Asp Arg Thr Met Lys Ser Ala  
 145 150 155 160  
 Gln Phe Phe Ala Lys Gly Ile Ala Leu Asp Ala Ser Lys Ile Ala Ile  
 165 170 175  
 Glu Arg Val Tyr Glu Gly Lys Lys Asp Gly Ala Asn Ser Leu Asn Pro  
 180 185 190  
 Tyr Asp Ser Cys Pro Ala Phe Ser Arg Val Thr Gly Ser Asp Glu Ala  
 195 200 205  
 Ser Val Phe Arg Glu Ile Tyr Thr Lys Pro Val Met Gln Arg Phe Asn  
 210 215 220  
 Ser Leu Ala Pro Ser Phe Asn Phe Thr Ser Thr Asp Ile Tyr Ala Phe  
 225 230 235 240  
 Ser Leu Ile Cys Gly Tyr Glu Thr Val Leu Arg Gly Ser Ser Pro Phe  
 245 250 255  
 Cys Gly Leu Ser Val Leu Ser Ser Asn Glu Trp Leu Gly Phe Glu Tyr  
 260 265 270  
 Ala Asn Asp Ile Lys Tyr Phe Tyr Asn Ser Gly Tyr Gly Leu Pro Ser  
 275 280 285  
 Ala Gly Ala Thr Gly Phe Pro Trp Val Asn Ala Ser Phe Asn Ala Leu  
 290 295 300  
 Met Ser Asn His Asp Thr Asn Ser Asn Glu Val Lys Asp Gln Asp Leu  
 305 310 315 320  
 Phe Val Ser Phe Thr His Arg Gly Met Pro Pro Met Ala Leu Val Ala  
 325 330 335  
 Met Gly Leu Phe Asn Asn Ser Ala Tyr Ser Gly Ala Asn Asn Ile Asn  
 340 345 350

Asp Thr Met Pro Leu Asp Thr Ile Asn Tyr Gln Arg Val Trp Lys Ser  
 355 360 365

Ser Gln Ile Ile Pro Phe Met Thr Asp Ile Gly Ile Glu Lys Leu Glu  
 370 375 380

Cys Asp Ser Phe Gly Phe Asp Glu Gly Ile Tyr Tyr Arg Val Leu Val  
 385 390 395 400

Asn Asp Arg Pro Gln Ser Leu Ile Gly Cys His Asp Gly Pro Ala Asp  
 405 410 415

Ser Cys Lys Glu Glu Ser Ile Arg Lys Trp Leu Asn Glu Arg Ala Lys  
 420 425 430

Val Val Gly Asp Phe Asp Thr Met Cys Arg Val Asp Tyr Ala Asn Ser  
 435 440 445

Thr Asn Ile Leu Ser Ile Tyr Asp Ser  
 450 455

&lt;210&gt; 13

&lt;211&gt; 1751

&lt;212&gt; DNA

&lt;213&gt; Triticum aestivum

&lt;400&gt; 13

gcacgagctt ttttaactcc cacaatatgc actctgttac attcttttct attctactga 60  
 ctggcgtggg cactaccaga gccgatgtgg actgatatc aacaagatat 120  
 ctaggaaattg gggacatctt agcccttac caacaatcg agatgatgac tttggtgtcg 180  
 aatatgtcgg acttcctgcc ggctgccaaa ttgaatctgc tcataccctc caacgccacg 240  
 cagaaagatt tccaaccaag gccataaacg atggcagaaa taatcaacgt ttacgtgaaa 300  
 aggtcacaaa ctctacttca acccatccaa aagatctgtt cactgggtccc ctccgattta 360  
 tgaattcgtg gaatccagtc atactcagta atggcctgtt aactggcatt ggtgcctcgg 420  
 ccgagtttaa tgctgggggt gagttctgga accggtacgg aagaacattg tacaatgcaa 480  
 gcgttggtga atagcctac aacggttcct atgctgatgg cacaccacgt ccgcccatag 540  
 ttctgaggac tactgagcaa tcgcggatgc ataaccacca gattaaactg gccctcggtc 600  
 ttctcgggcc cagcttttca ccgtaacca atccagctct aaatgacacg gctaaagcat 660  
 tcgaggtagt cattatccca gaagaaaatg gcggcaaa caataataca ctggcttcat 720  
 atgtaagctg ttcaaatgct ggaaccgcga ctatgaaggc ttccaccaac cagatgcttt 780  
 atgagtttgt ctctagttaa ctagggtccc ccaccgaaag actcaaatct tttgtaccag 840  
 cagattttcc actgacgctt aatgacacat ttgccatgca gatgatttgt gcctacgaga 900  
 atgcactcat gggagcatca gaattttgtg gtttttttac agaagatgag tggactggat 960  
 ttgagaacag cctagaaatg caatttctat acaaattttc ttacggtcat cccactggtc 1020  
 gcgctcaagg tatcgctat gtcgaggaac ttttggcccg cctaaataac acccagatga 1080  
 caagctcttc tacttcagtt aatgcaacgc taaatagcga cccaaaaaac ttccagataa 1140  
 accaaacott ttatgcgcac ttactcatg acactgttat cacatcggtt ctctccgctt 1200  
 gtcaataga ctatttccat gcacctccta atttaaccga aatctcatcc gaocaaaaac 1260  
 gtaatttat actctcgaaa attgtgcctt ttggcgccg ctagttacc gagacaatcg 1320  
 gctgcacccc acccaaccga aagccgcgca agactgttat agtatgaga acacctgaac 1380  
 aatacggctc cgagaaatct ttacgtaaat acaaatttgt acgtatgaga ttaaatatg 1440  
 gcatatatac attagcacaca atccgtggag gcgctgtgtg tgatgacaag agcagccgaa 1500  
 ttgacgggct ttgcgctttg gatgatttta tgacaagcca cgagaatgcc tggccatgg 1560  
 cgaattatca atatgcttgt tacgggaatt atactaacca gggccgggcc actgggtggg 1620  
 attgggatgg aactatttca gaatagctgg accatgtttc ggatttcaaa ccccgctgta 1680  
 attgctgtat actacgagag taatttttaa atgtaaaaga tgatggacac aaaaaaaaaa 1740  
 aaaaaaaaaa a 1751

<210> 14  
 <211> 539  
 <212> PRT  
 <213> Triticum aestivum

<400> 14

```

Met His Ser Val Thr Phe Phe Ser Ile Leu Leu Thr Gly Val Val Thr
 1           5           10           15

Thr Arg Ala Asp Val Ala Arg Val Val Thr Asp Ile Asn Lys Ile Ser
          20           25           30

Arg Asn Trp Gly His Leu Ser Pro Tyr Ala Asn Asn Arg Asp Asp Asp
 35           40           45

Phe Gly Val Glu Tyr Val Gly Leu Pro Ala Gly Cys Gln Ile Glu Ser
 50           55           60

Ala His Thr Leu Gln Arg His Ala Glu Arg Phe Pro Thr Lys Gly Ile
 65           70           75           80

Asn Asp Gly Arg Asn Asn Gln Arg Phe Ser Glu Lys Val Thr Asn Phe
          85           90           95

Thr Ser Thr His Pro Lys Asp Leu Phe Thr Gly Pro Leu Arg Phe Met
100           105           110

Asn Ser Trp Asn Pro Val Ile Leu Ser Asn Gly Leu Leu Thr Gly Ile
115           120           125

Gly Ala Ser Ala Glu Phe Asn Ala Gly Val Glu Phe Trp Asn Arg Tyr
130           135           140

Gly Arg Thr Leu Tyr Asn Ala Ser Val Gly Gln Leu Ala Tyr Asn Gly
145           150           155           160

Ser Tyr Ala Asp Gly Thr Pro Arg Pro Pro Ile Val Leu Arg Thr Thr
165           170           175

Glu Gln Ser Arg Met His Asn Thr Gln Ile Asn Trp Ala Leu Gly Phe
180           185           190

Phe Gly Pro Ser Phe Ser Pro Val Pro Asn Pro Ala Leu Asn Asp Thr
195           200           205

Ala Lys Ala Phe Glu Val Val Ile Ile Pro Glu Glu Asn Gly Gly Lys
210           215           220

Gln Asn Asn Thr Leu Ala Ser Tyr Val Ser Cys Ser Asn Cys Gly Asn
225           230           235           240

Pro Thr Met Lys Ala Ser Thr Asn Gln Met Leu Tyr Glu Phe Val Ser
245           250           255

Ser Tyr Leu Gly Pro Ala Thr Glu Arg Leu Lys Ser Phe Val Pro Ala
260           265           270

Asp Phe Pro Leu Thr Val Asn Asp Thr Phe Ala Met Gln Met Ile Cys
275           280           285
  
```

Ala Tyr Glu Asn Ala Leu Met Gly Ala Ser Glu Phe Cys Gly Phe Phe  
 290 295 300  
 Thr Glu Asp Glu Trp Thr Gly Phe Glu Asn Ser Leu Glu Met Gln Phe  
 305 310 315  
 Tyr Tyr Lys Phe Ser Tyr Gly His Pro Thr Gly Arg Ala Gln Gly Ile  
 325 330 335  
 Gly Tyr Val Glu Glu Leu Leu Ala Arg Leu Asn Asn Thr Gln Met Thr  
 340 345 350  
 Ser Ser Ser Thr Ser Val Asn Ala Thr Leu Asn Ser Asp Pro Lys Thr  
 355 360 365  
 Phe Pro Val Asn Gln Thr Phe Tyr Ala Asp Phe Thr His Asp Thr Val  
 370 375 380  
 Ile Thr Ser Val Leu Ser Ala Leu Ser Ile Asp Tyr Phe His Ala Pro  
 385 390 395 400  
 Pro Asn Leu Thr Gln Ile Ser Ser Asp Pro Asn Arg Lys Phe Ile Leu  
 405 410 415  
 Ser Lys Ile Val Pro Phe Gly Ser Arg Leu Val Thr Glu Thr Ile Gly  
 420 425 430  
 Cys Thr Ser Pro Asn Pro Lys Pro Arg Lys Ser Ala Gln Val Gln Tyr  
 435 440 445  
 Thr Pro Glu Gln Tyr Gly Tyr Glu Lys Ser Leu Ala Lys Tyr Lys Phe  
 450 455 460  
 Val Arg Met Arg Leu Asn Asn Gly Ile Ile Pro Leu Asp Thr Ile Arg  
 465 470 475 480  
 Gly Gly Ala Cys Gly Asp Asp Lys Ser Ser Arg Ile Asp Gly Leu Cys  
 485 490 495  
 Ala Leu Asp Asp Phe Met Thr Ser Gln Gln Asn Ala Ser Ala Met Ala  
 500 505 510  
 Asn Tyr Gln Tyr Ala Cys Tyr Gly Asn Tyr Thr Asn Gln Gly Arg Ala  
 515 520 525  
 Thr Gly Trp Asp Trp Asp Gly Thr Ile Ser Glu  
 530 535

<210> 15  
 <211> 507  
 <212> DNA  
 <213> Triticum aestivum

<220>  
 <221> unsure  
 <222> (307)

<220>  
 <221> unsure  
 <222> (435)

<220>  
 <221> unsure  
 <222> (500)..(501)..(502)

<400> 15  
 gtactcaata agagcaaccc aggttctctg agcatcagct agtgacagtag catttggttt 60  
 ggggctactt tctgggaaag gaaagcttgg agcaggaat aaccgtgcct tctctgttct 120  
 gaggtagagt cgtgcaagtg atatttgtct gcgattcttt gatagctgta aaacatacaa 180  
 ggactacagg aaaagaaaagg agcctgatgt tgacaagcaa aaggaaccaa ttctaagaac 240  
 atgtcacatc tgctttagtc agccgttatc accccaagtt tacaacacag gatgtttctt 300  
 ccctccnggt tctcttgcaa acaggaagca tctttgctga atattaccaa tcaagcttgt 360  
 caacttttca atgaagatga ggtcaattgc tagatggagc agatgatttg gaggggttgg 420  
 gcctaaaggt atgngagca ataaacttaa aatgggactg caatgctcaa ggagttgtca 480  
 atccatgaa gaacatcatn nnagaga 507

<210> 16  
 <211> 129  
 <212> PRT  
 <213> Triticum aestivum

<400> 16  
 Tyr Ser Ile Arg Ala Thr Gln Val Pro Arg Ala Ser Ala Ser Ala Val  
 1 5 10 15  
 Ala Phe Gly Leu Gly Leu Leu Ser Gly Lys Gly Lys Leu Gly Ala Gly  
 20 25 30  
 Asn Asn Arg Ala Phe Ser Val Leu Ser Glu Ser Arg Ala Ser Asp Ile  
 35 40 45  
 Cys Leu Arg Phe Phe Asp Ser Cys Lys Thr Tyr Lys Asp Tyr Arg Lys  
 50 55 60  
 Arg Lys Glu Pro Asp Val Asp Lys Gln Lys Glu Pro Ile Leu Arg His  
 65 70 75 80  
 Val Thr Ser Ala Leu Val Ser Arg Tyr His Leu Lys Phe Thr Thr Gln  
 85 90 95  
 Asp Val Ser Ser Leu Arg Phe Leu Cys Lys Gln Glu Ala Ser Leu Leu  
 100 105 110  
 Asn Ile Thr Asn Gln Ala Cys Gln Leu Phe Asn Glu Asp Glu Val Asn  
 115 120 125  
 Cys

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
18 January 2001 (18.01.2001)

PCT

(10) International Publication Number  
WO 01/04147 A3

- (51) International Patent Classification: C12N 15/82, 15/55, 9/16, 5/10, C12Q 1/68, A01H 5/00
- (21) International Application Number: PCT/US00/18778
- (22) International Filing Date: 11 July 2000 (11.07.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/143,411 12 July 1999 (12.07.1999) US
- (71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).
- (72) Inventor: and
- (75) Inventor/Applicant (for US only): CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US).
- (74) Agent: GEIGER, Kathleen, W.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).
- (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:  
23 August 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/04147 A3

(54) Title: PLANT INOSITOL POLYPHOSPHATE PHOSPHATASE HOMOLOGS

(57) Abstract: This invention relates to an isolated nucleic acid fragment encoding an inositol polyphosphate phosphatase. The invention also relates to the construction of a chimeric gene encoding all or a substantial portion of the inositol polyphosphate phosphatase, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the inositol polyphosphate phosphatase in a transformed host cell.